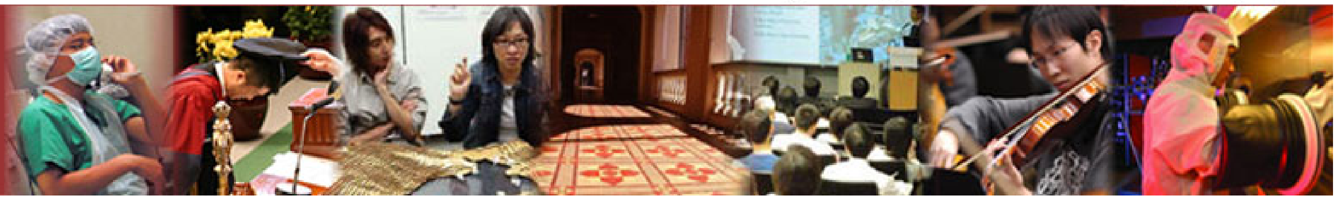

Title	Acoustic and electroglottographic (EGG) characteristics of tracheoesophageal speech of Cantonese
Author(s)	Chiu, Ka-lai; 趙嘉麗
Citation	Chiu, K. [趙嘉麗]. (2012). Acoustic and electroglottographic (EGG) characteristics of tracheoesophageal speech of Cantonese. (Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR.
Issued Date	2012
URL	http://hdl.handle.net/10722/237897
Rights	The author retains all proprietary rights, (such as patent rights) and the right to use in future works.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Acoustic and electroglottographic (EGG) characteristics of tracheoesophageal speech of Cantonese

CHIU, Ka Lai Nana

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2012.

Abstract

This study investigated the acoustic and electroglottographic (EGG) characteristics of tracheoesophageal (TE) speech of Cantonese. Seven TE speakers and seven aged-matched laryngeal (NL) speakers produced sustained vowel phonation and passage reading. Both acoustic and EGG signals were recorded and analyzed using software programs Praat and Voce Vista. Results indicated that there was a significantly lower fundamental frequency (F0) for TE than NL speakers in passage reading. Significantly higher jitter, shimmer and closed quotient (CQ), and lower harmonic-to-noise ratio (H/N) values were associated with TE speech in both sustained vowel phonation and passage reading. Higher formant frequencies in sustained /i/ phonation were found for TE speakers. The findings appear to explain the perceptually hoarse, breathy and low-pitch voice of TE speech. Results were discussed in terms of higher position, greater tissue density, slower movement during closing phase and aperiodic vibration of neoglottis of TE speakers than the vocal folds of NL speakers.

Introduction

Total laryngectomy is the removal of a pathological larynx commonly as a result of laryngeal cancer. During the surgery, the entire larynx is removed, and the upper end of the trachea is attached to an opening known as stoma created at the anterior neck, leaving the upper pulmonary airway unprotected (Stemple, Glaze, & Klaben, 2010). It is through the stoma the laryngectomee breathes. After the procedure, laryngectomees lose their ability to make sound, as in laryngeal speakers. Finding an alternative way to talk is thus a vital issue in the post-surgery rehabilitation.

To regain phonation, a new sound source is adopted by laryngectomees in producing different alaryngeal speech sounds. Currently, there are four types of alaryngeal speech being used by laryngectomees in Hong Kong. They are esophageal (SE) speech, tracheoesophageal (TE) speech, pneumatic artificial laryngeal (PA) speech, and electrolaryngeal (EL) speech (Ng, Kwok, & Chow, 1997).

Tracheoesophageal (TE) speech

To facilitate TE phonation, a voice prosthesis is inserted in a fistula that is surgically created between the posterior tracheal wall and anterior esophageal wall (Singer & Blom, 1980). The voice prosthesis is a one-way valve which prevents aspiration during swallowing and allows pulmonary air to pass through and enter esophagus by occluding the stoma during exhalation. Successful entry of air from the trachea into the esophagus provides the air reservoir for sound production. As air pressure built up below the pharyngoesophageal (PE) segment becomes sufficiently high, the PE segment will be set into vibration. The vibration thus creates sound waves which travel along the vocal tract for speech sound production. The PE segment comprises of cricopharyngeus muscle of the upper esophagus and muscles of the middle and inferior pharyngeal constrictor (Stemple, Glaze, & Klaben, 2010). It substitutes the role of vocal cords in laryngeal speakers and serves as the source of sound in alaryngeal speech. It is therefore called the neoglottis.

Acoustic Analysis

Previous studies have investigated various acoustic characteristics associated with TE voice. Ng and Wong (2009) examined the voice onset time (VOT) in aspirated (/p^h/, /t^h/, & /k^h/) and unaspirated stops (/p/, /t/, & /k/) produced by TE speakers, SE speakers and laryngeal speakers of Cantonese. Their results revealed that all the three speaker groups demonstrated significantly longer VOT in aspirated stops than in unaspirated stops, indicating that TE speakers could still demonstrate the distinction for aspiration, despite the use of neoglottis as the new sound source.

Apart from VOT measurements, formant frequencies associated with vowels produced by alaryngeal speakers have also been investigated (Liu & Ng, 2009; Ng & Chu, 2009; Sisty & Weinberg, 1972). Ng and Chu (2009) found significantly higher formant frequencies associated with TE speech than laryngeal speakers of Cantonese. The finding suggested that TE speakers predominately spoke with a PE segment located at a higher location than laryngeal speakers, implying a reduction of effective vocal tract length for resonance of speech sound after laryngectomy. Liu and Ng (2009) examined the formant characteristics of SE speakers of Mandarin in a vowel production task. The study also revealed significantly higher formant frequencies (F1, F2, & F3) in SE speakers than laryngeal speakers of Mandarin. The implication of a shorter vocal tract for resonance with higher PE segment position was thus further supported. The findings proved that the ‘filter’ of sound production is also altered in TE and SE speakers, compared with laryngeal speakers. Similar findings were reported by Sisty and Weinberg (1972) based on data obtained from English-speaking laryngectomees.

There are also studies investigating the fundamental frequency associated with alaryngeal speech. According to Ng, Gilbert, and Lerman (2001), fundamental frequency contours were found to be acoustic cues for the perception of esophageal speech of Cantonese. Another study comparing fundamental frequency between laryngeal speech and alaryngeal speech of Hebrew speakers revealed significantly lower fundamental frequency for both TE and SE speech than laryngeal speakers (Most, Tobin, & Mimran 2000).

Moreover, TE speakers are generally thought to have reduced pitch, pitch variation, intensity, intensity range, and speech rate and greater jitter, shimmer and harmonic-to-noise ratio (H/N). The aperiodicity and irregularity of TE and esophageal speech were explained in terms of poor voluntary control of the PE segment by Robbins, Fisher, Blom, & Mark (1984).

Electroglottography (EGG)

Electroglottography provides a non-invasive way to examine vocal fold movements during phonation (Kitzing, 2000). Fabre (1957) was the pioneer in using high frequency glottography to investigate the vibratory behavior of vocal folds during phonation. Two metal electrodes are placed externally on the two sides of the thyroid cartilage at the level of the glottis. A high frequency current is passed through the electrodes. Impedance of air is considered to be infinitely high while that of body tissue and fluid are good conductors with low impedance. During the closing phase of a glottal cycle, the vocal folds approximate each other and area of contact increases. Thus, the electric impedance between the two electrodes decreases and the electric current increases. The EGG waveform reflects the amount of transverse impedance at the laryngeal level which varies inversely with the relative amount of vocal fold contact during phonation (Rothenberg & Mahshie, 1988).

EGG has been used in measuring vocal fold vibration in normal adults (Childers, Krishnamurthy, Naik, Larar, & Moore, 1983; Ma & Love, 2010), and pathological adult population (Childers, Smith & Moore, 1984; Konstantopoulos, Vikelis, Seikel, & Mitsikostas, 2010). Ma and Love (2010) documented the use of mean F0 and mean contact quotient obtained from EGG signals to examine the age and gender effect in sustained phonation and connected speech tasks. Mean F0 values of females were found significantly higher than that of males. It also revealed significantly lowered mean F0 for the aged females than the young females. The study also found significantly smaller contact quotient in older males than in young males, and a significantly larger contact quotient in older females than in young females. The discrepancy of the mean F0 and contact quotient was attributed to the anatomical changes to the laryngeal muscles in normal aging.

Konstantopoulos et al. (2010) compared the EGG and acoustic parameters in speakers with multiple sclerosis (MS) and the normal speakers. Results demonstrated a significantly greater standard deviation of fundamental frequency and jitter, and a significantly lower fundamental frequency and range of fundamental frequency in the MS group than the normal group. The authors attributed the difference to the symptoms of the disease, i.e. incoordination, reduction in range and force of movement during vocal fold adduction.

Recently, EGG has been used as an objective assessment tool for TE speech research (Kazi et al., 2006; 2009a). In Kazi et al. (2009a), evaluation of PE segment vibratory characteristics using EGG was performed and it was validated to assess TE speech for clinical use. The study investigated PE segment vibratory characteristics in terms of different voice parameters during production of vowel and connected speech. Result showed that there were five parameters including F0, jitter, shimmer, normalized noise energy (NNE), and irregularity correlated well with the perceptual assessment. Jitter, shimmer and irregularity significantly correlated with the overall grade of hoarseness and overall voice quality of TE speech. Average F0, shimmer and NNE also significantly correlated with breathiness of the TE speech.

There are insufficient local studies documenting the objective measures of different acoustic and EGG parameters of TE speakers of Cantonese. Only voice onset time (Ng & Wong, 2009) and formant frequencies (Ng & Chu, 2009) had been documented for TE speech of Cantonese. Other acoustic parameters including F0, standard deviation of fundamental frequency, jitter, shimmer and HNR were found to be significantly correlated with TE speech but were documented in foreign studies only. There is also a lack of documentary of EGG parameters of TE speakers of Cantonese. It was therefore the aim of the present study to investigate the objective voice characteristics associated with Cantonese TE speakers in terms of acoustic and EGG analyses in both sustained phonation and connected speech. The study was set out to evaluate the significance of acoustic parameters and EGG parameters in differentiating TE speech from laryngeal speech of Cantonese and suggest possible explanation to the differences.

Clinical application

By examining the voice characteristics associated with TE speech using acoustic and EGG analyses, valuable references for supplementing perceptual evaluation of TE speech can be achieved. The combined use of both objective measures and perceptual evaluation could provide a multi-dimensional view of TE speech, as voice quality is after all a multi-dimension entity. This information can assist clinicians to assess the competence of TE speech production which may result in application of different vocal rehabilitation strategy by the professional.

For vocal rehabilitation, the use of objective measures could possibly identify the area of deficiencies in TE speaker of which clinicians should target on as one of the treatment goal. Besides, clinicians could make use of the objective measures e.g. EGG as a visual feedback in order to facilitate vocal rehabilitation. For instance, EGG allows a display of the electric current flow through the instrument during opening and closing phases of neoglottal vibration. In addition, results of objective analysis of TE speech provide more realistic information about TE speakers’ communication function. Such information will then be used to refine existing post-laryngectomy speech rehabilitation for various alaryngeal speakers.

Methods

Participants

Seven TE speakers were carefully selected from the New Voice Club of Hong Kong, which is a non-profit self-help organization for laryngectomees in Hong Kong. They were all adult males aged from 54 to 85 years (mean age = 66.14 years, SD = 11.6 years). All alaryngeal speakers were physically healthy and judged as superior TE speakers by practicing speech-language pathologists. Information of age, duration of using TE speech and use of other alternative alaryngeal speech of the TE speakers were presented in Table 1. Seven laryngeal speakers selected from the family and friends of the investigator were invited to participate in the study. The laryngeal speakers were all adult males and aged from 50 to 79 years (mean age = 66.57 years, SD = 10.6 years). All participants were native speakers of Cantonese with no known history of speech problems. All participants were able to read speech materials used in the study.

Table 1. Age, duration of experience in tracheoesophageal (TE) speech and use of other alaryngeal speech in TE speakers.

Subjects	Age (yrs)	Duration of experience in TE speech (yrs)	Use of other alaryngeal speech		
			SE	PA	EL
1	54	5	yes	no	no
2	85	1	no	no	no
3	80	11	yes	yes	yes
4	59	6	yes	no	no
5	63	3	no	no	no
6	60	1	no	no	no
7	62	1	yes	no	no

Notes. SE = standard esophageal speech. PA = pneumatic artificial laryngeal speech. EL = electrolaryngeal speech.

Instrumentation and Procedures

Acoustic and EGG signals were obtained using the two-channel Electroglottograph and Microphone Amplifier; model EG2-PCX2 (Glottal Enterprises, Inc.) With the participant comfortably seated in a sound treated room, a pair of electrodes (gold-plated, diameter 34 mm) cleaned with alcohol swab, were placed externally on each side of the thyroid lamina in normal speakers or a few centimeters below the stoma and 3 - 4 centimeters apart in TE speakers to detect EGG signals during voicing. Sound was recorded by a headset electret capacitor microphone (Glottal Enterprise M-80) held at approximately 4 centimeters from the corner of the mouth. The acoustic signal was recorded simultaneously with the EGG signal using a high-quality microphone plugged into the front panel of the EGG device. A computer equipped with a sound card was connected to the EGG device via a stereo cable whereas the left and right channels carried the acoustic and EGG signals respectively. The signals were sampled at frequency 22 kHz and recorded using the Voce Vista Signals (Glottal Enterprises, Inc.).

Speech tasks

Two speech tasks were administered in the present study. The first task was to obtain acoustic and EGG information from sustained vowel production (/a/, /i/, and /u/). Each participant was required to sustain a vowel at a time for three seconds at his comfortable pitch and loudness. This task was repeated three times.

The second task was to obtain acoustic and EGG information in connected speech context. Each participant was required to read the passage of ‘Barbara Stuart’ at his comfortable pitch, loudness, and pace. A brief period of time was given for the participants to practice and get familiarized with the passage content before actual recording.

Measures

Among the three recordings of each sustained vowel production, the best attempt was selected and only the middle portion of the recording was used for acoustic and EGG analysis. This was judged by the investigator who was experienced in analyzing TE speech. For the passage task, the whole recording of the passage was used for analysis.

Acoustic measures

Seven voice parameters were obtained by analyzing the acoustic waveforms using software program Praat (Boersma, 2001). They included: (1) mean fundamental frequency (F0), (2) standard deviation of F0 (SD-F0), (3) first formant (F1), (4) second formant (F2), (5) percent jitter, (6) shimmer in dB, and (7) harmonic-to-noise ratio (H/N). For measuring F0, the method of cross-correlation was used. Default settings in the Praat were kept, except for pitch extraction range, which was adjusted as from 20 Hz to 500 Hz. For percent jitter measurement, it was calculated by dividing the absolute difference between consecutive periods by the average period and was expressed in percentage. Shimmer in dB was obtained by finding the average value of absolute base-10 logarithm of the difference between amplitudes of two consecutive periods, and then multiplying by 20, expressed in decibel. Finally, H/N ratio is a measure of harmonicity or periodicity, reflecting the signal-to-noise ratio of the source of vibration which generates the acoustic signal. It was calculated by taking base-10 logarithm ratio of energy in the periodic signal to that in the noise, and then multiplying 10, expressed in decibel.

EGG measures

Three voice parameters were obtained from analysis of EGG signal. Mean F0 and its standard deviation (SD-F0) were analyzed using Praat while closed quotient (CQ) was calculated using the software Voce Vista (Glottal Enterprises, Inc.). CQ is the ratio of closed time to period in one complete glottal cycle. In a EGG waveform, upward slope is regarded as closing of the vocal folds (in normal speakers) or PE segment (in TE speakers) while downward slope as opening of vocal folds (in normal speakers) or PE segment (in TE speakers). The closing moment was identified as the point of greatest rising slope whereas the opening moment was identified as the point of greatest falling slope. The time between the two moments are the closed time while the time between two consecutive closing moments was the period. Using Voce Vista, after identification of the two moments, CQ is automatically calculated and displayed (see Figures 1 and 2).

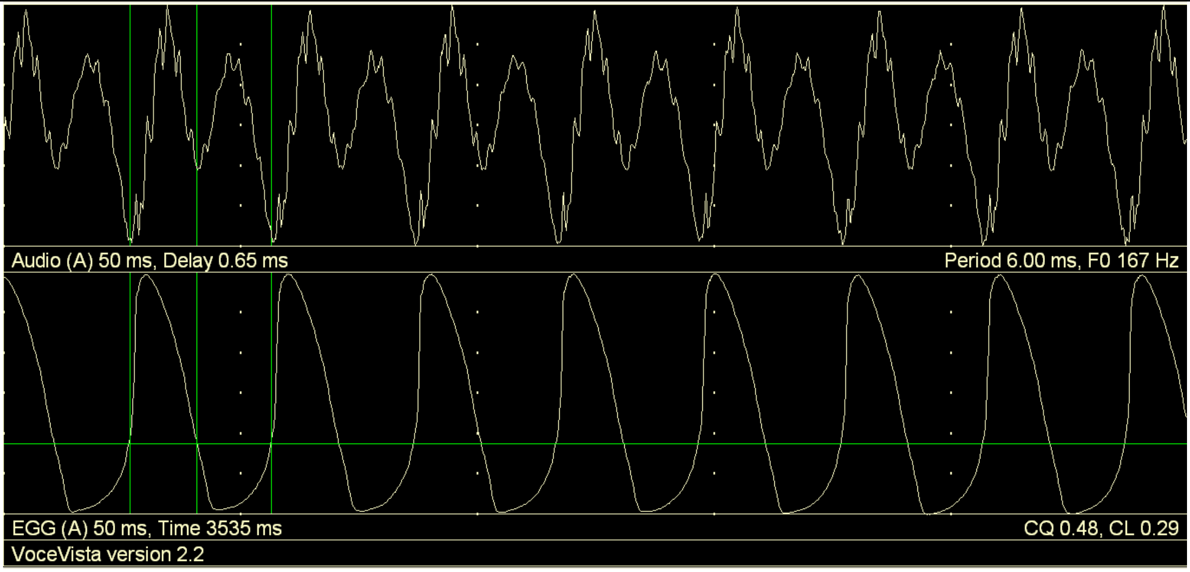

Figure 1. Synchronous acoustic and electroglottographic waveforms of sustained /i/ production of a laryngeal speaker.

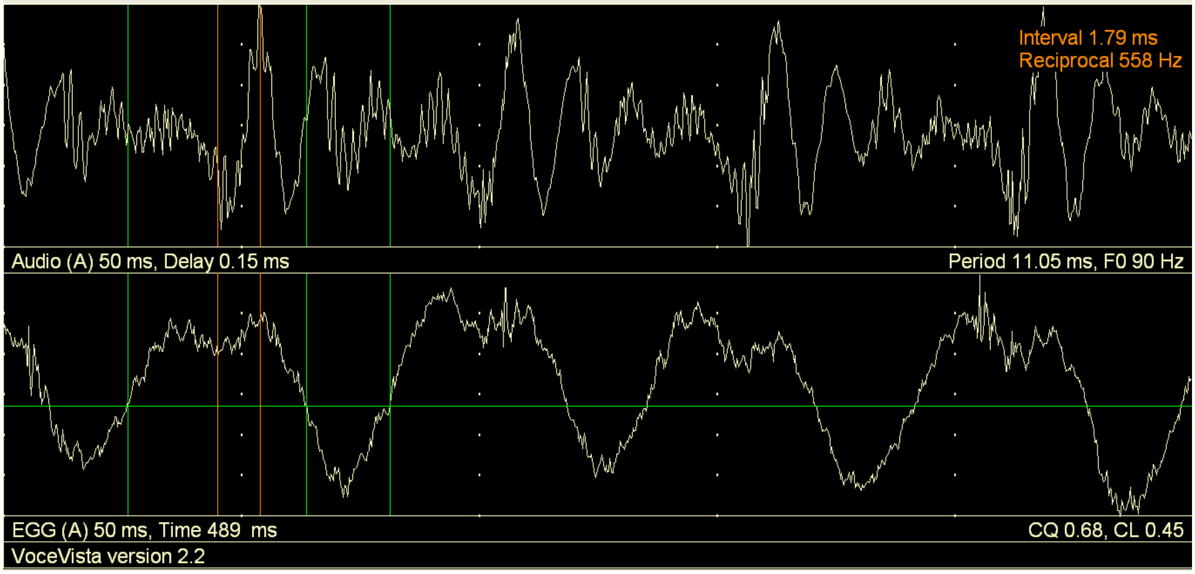

Figure 2. Synchronous acoustic and electroglottographic waveforms of sustained /i/ production by a tracheosophageal speaker.

Temporal measure

Apart from the above parameters, another parameter word per minute (WPM) was considered as a temporal measure to estimate the fluency of the speech of the participants. WPM was calculated from the total time used for oral reading the passage 'Barbara Stuart' which contained a total number of 123 words.

Data Analysis

Data from laryngeal and TE speaker groups were compared using independent-samples t-tests. Levene's test for equality of variances was used to determine whether equal variance was assumed.

Results

Sustained vowel analysis

Comparison of mean and range values of various voice parameters between laryngeal group and TE group during sustained production of vowels /a/, /i/ and /u/ are presented in Tables 2 to 4 correspondingly. According to the Tables, it is noted that TE group was characterized with lower mean pitch (except higher pitch in sustained vowel /a/ production) and mean fundamental frequency than the normal group in sustained production of all the three vowels. The TE group also demonstrated significantly higher SD-pitch and SD-F0 than the laryngeal group during sustained vowel /i/ production ($p < .05$). In formant analysis, the TE group was found to have significantly ($p < .05$) higher F1 (except not significant in sustained /a/) and F2 (except not significant in sustained /u/) than the normal group. In addition, significantly greater jitter, shimmer, and closed quotient values, and lower HNR values were found for TE speakers when compared with laryngeal speakers ($p < .05$).

Connected speech analysis

Comparison of mean and range values of various voice parameters between laryngeal group and TE group in connected speech production are presented in Table 5. Similar to sustained vowels, the TE group had greater SD-pitch and F0, F1, F2, jitter, shimmer, and CQ, and lower HNR values than the laryngeal group. All the differences reached statistical significance ($p < .05$), except the standard deviation of pitch.

The mean pitch and mean F0 of the TE group was found larger than the laryngeal group with mean F0 reached statistical significance ($p < .05$). Smaller WPM value in TE speakers than laryngeal speakers was found. However, the difference did not reach statistical significance ($p = .178$).

Table 2. Mean (and range) of values voice parameters of laryngeal group and tracheoesophageal (TE) group in sustained /a/ phonation.

Voice parameters	NL Group (N=7)	TE Group (N=7)	*p* value
Pitch (Hz)	138.6 (126.1-149.9)	249.9 (30.9-422.9)	0.121
SD-Pitch (Hz)	1.36 (0.84-2.2)	30.4 (0.79-124.4)	0.149
F0 (Hz)	138.5 (126.1-149.9)	104.8 (31.2-207.7)	0.255
SD-F0 (Hz)	1.36 (0.84-2.2)	32.8 (0.91-121.5)	0.079
F1 (Hz)	697.0 (576.2-843.2)	784.3 (590.2-922.4)	0.109
F2 (Hz)	1162 (978-1311)	1405 (1247-1645)	0.007*
Jitter (%)	0.35 (0.21-0.69)	4.44 (0.43-8.32)	0.012*
Shimmer (dB)	0.35 (0.17-0.80)	1.61 (0.70-2.26)	0.001*
H/N (dB)	18.8 (14.6-21.5)	2.01 (0.51-5.60)	0.000*
Closed Quotient	0.42 (0.29-0.54)	0.57 (0.35-0.76)	0.029*

Note. SD = standard deviation, F0 = fundamental frequency, F1 = 1st formant frequency, F2 = 2nd formant frequency, H/N = harmonic-to-noise-ratio.

* $p < 0.05$.

Table 3. Mean (and range) values of voice parameters of laryngeal group and tracheoesophageal (TE) group in sustained /i/ phonation.

Voice parameters	NL Group (N=7)	TE Group (N=7)	*p* value
Pitch (Hz)	146.0 (121.3-171.5)	123.2 (31.2-230.5)	0.462
SD-Pitch (Hz)	1.44 (0.99-2.1)	9.54 (1.59-23.1)	0.034*
F0 (Hz)	146.1 (121.3-171.5)	124.4 (62.7-245.4)	0.421
SD-F0 (Hz)	1.44 (0.99-2.16)	30.1 (1.67-83.0)	0.042*
F1 (Hz)	302.0 (265.5-339.5)	654.1 (290.9-1259)	0.048*
F2 (Hz)	2162 (1824-2357)	2544 (2255-2889)	0.005*
Jitter (%)	0.25 (0.14-0.46)	2.99 (0.77-6.55)	0.01*
Shimmer (dB)	0.16 (0.07-0.30)	1.08 (0.52-1.47)	0.000*
H/N (dB)	22.6 (20.2-28.0)	5.06 (0.49-12.1)	0.000*
Closed Quotient	0.47 (0.38-0.55)	0.58 (0.46-0.68)	0.014*

Note. SD = standard deviation, F0 = fundamental frequency, F1 = 1st formant frequency, F2 = 2nd formant frequency, H/N = harmonic-to-noise-ratio.

* $p < 0.05$.

Table 4. Mean (and range) values of voice parameters of laryngeal group and tracheoesophageal (TE) group in sustained /u/ phonation.

Voice parameters	NL Group (N=7)	TE Group (N=7)	*p* value
Pitch (Hz)	149.5 (136.5-180.0)	113.2 (66.7-205.8)	0.093
SD-Pitch (Hz)	1.43 (0.80-2.28)	18.6 (2.29-54.1)	0.051
F0 (Hz)	149.5 (136.5-180.0)	124.0 (80.9-208.4)	0.203
SD-F0 (Hz)	1.42 (0.80-2.27)	18.4 (2.52-49.1)	0.069
F1 (Hz)	435.5 (329.2-531.2)	545.1 (440.4-655.3)	0.019*
F2 (Hz)	1077 (656.1-2332)	1497 (1185-2146)	0.131
Jitter (%)	0.26 (0.12-0.49)	2.60 (0.61-4.03)	0.000*
Shimmer (dB)	0.18 (0.07-0.28)	1.22 (0.53-1.74)	0.000*
H/N (dB)	26.3 (20.4-31.4)	7.52 (3.15-14.89)	0.000*
Closed Quotient	0.40 (0.33-0.51)	0.53(0.48-0.72)	0.011*

Note. SD = standard deviation, F0 = fundamental frequency, F1 = 1st formant frequency, F2 = 2nd formant frequency, H/N = harmonic-to-noise-ratio.

* $p < 0.05$.

Table 5. Mean (and range) values of voice parameters of laryngeal group and tracheoesophageal (TE) group in connected speech production.

Voice parameters	NL Group (N=7)	TE Group (N=7)	*p* value
Pitch (Hz)	117.2 (99.8-132.9)	128.0 (92.1-184.6)	0.421
SD-Pitch (Hz)	36.1 (28.9-52.1)	77.2 (37.3-170.0)	0.073
F0 (Hz)	130.3 (92.6-193.4)	176.6 (121.9-243.8)	0.045*
SD-F0 (Hz)	52.6 (23.3-104.5)	135.0 (80.5-184.0)	0.001*
F1 (Hz)	616.8 (534.8-682.4)	729.4 (609.6-788.9)	0.002*
F2 (Hz)	1747(1632-1795)	1857 (1750-1968)	0.012*
Jitter (%)	2.37 (2.02-2.80)	3.77 (2.57-5.54)	0.004*
Shimmer (dB)	0.96 (0.83-1.10)	1.43 (1.17-1.79)	0.000*
H/N (dB)	8.62 (6.69-9.90)	3.67 (2.25-5.51)	0.000*
Closed Quotient	0.38 (0.25-0.55)	0.51 (0.42-0.66)	0.023*
Word Per Minute	175 (152-192)	151 (109-225)	0.178

Note. SD = standard deviation, F0 = fundamental frequency, F1 = 1st formant frequency, F2 = 2nd formant frequency, and H/N = harmonic-to-noise-ratio.

* $p < 0.05$.

Discussion

The results of this study demonstrated significant deviation of values for voice parameters between TE speakers and laryngeal speakers of Cantonese by both acoustic and EGG analysis in sustained vowel phonation and connected speech production. Apart from documenting the voice differences between the TE and laryngeal speakers of Cantonese and discussing their significances, a question of interest is why there was a difference. Comparisons between the results of this study and the literature, and subsequent discussion could establish significance of the findings and plausible explanations.

Pitch and low voice

Results obtained from acoustic analysis of pitch in sustained phonation of /i/ and /u/ revealed lower pitch in TE speakers than laryngeal speakers of Cantonese. This finding agrees with Most, Tobin, and Mimran (2000), in which acoustic analysis of vowel production demonstrated that TE speakers had lower fundamental frequency than laryngeal speakers of Hebrew. This consistent finding across languages supports the general perception of lower voice in TE speakers than laryngeal speakers. On the other hand, there was an exception in sustained /a/ phonation, in which the acoustic analysis showed a higher pitch in TE speakers than laryngeal speakers. This could be due to the effect of extreme pitch value in one of the TE subject (pitch = 422.9 Hz, SD = 124.4 Hz).

Aperiodicity and Voice Quality

TE speech is always thought to be aperiodic and perceived to be hoarse and breathy, compared with laryngeal speech. In this study, SD-Pitch, SD-F0, jitter, shimmer, and H/R of TE speech were compared with laryngeal speech of Cantonese. In both sustained vowel phonation and connected speech production, TE speakers had significantly higher SD-Pitch, SD-F0, jitter and shimmer and lower H/R than laryngeal speakers. A significantly greater jitter was noted in TE speakers which correlated well with perceptual assessment of overall grade of hoarseness and overall voice quality in Kazi et al. (2009a). In Ng, Kwok, and Chow (1997), the hoarse voice quality of TE speakers was found significantly poorer than PA speakers who employed an external device to produce a stable and periodic source of vibration. The PE segment in TE speakers was hard to be voluntarily controlled and thus more aperiodic vibration will be expected. Greater SD-pitch, SD-F0, jitter and noise were thus resulted from aperiodic vibration of the sound source in TE speakers. Shimmer was also found to be significantly higher in TE speakers. In Kazi et al. (2009a), shimmer correlated well with perceptual evaluation of breathiness, overall grade of hoarseness and overall voice quality in TE speakers. Therefore, it could be possible to infer TE speech as breathy and hoarse due to the aperiodic vibration of the sound source, according to the review of the literatures.

Formant frequencies and Length of vocal tract

Formant pattern depends on the vocal tract configuration. Changes of the configuration should therefore result in an alternation of the formant frequencies. In this study, significantly higher formant frequencies were found in TE speakers than laryngeal speakers, with the exception of F1 of sustained /a/ phonation and F2 of sustained /u/ phonation. Higher formant frequencies imply a shorter effective length of vocal tract for resonance during phonation in TE speakers. The phenomenon was coherent with the studies done by Ng and Chu (2009), and Liu and Ng (2009). After total laryngectomy, a new sound source, i.e. the PE segment, substituted the vocal folds for the vibration during phonation. According to a videofluoroscopic examination of the neoglottis by van As, Op de Coul, van den Hoogen, Koopmans-van Beinum, and Hilgers (2001), TE speakers were found to have neoglottis at position of cervical vertebrae C4 to C5 at rest and most speakers reached a higher neoglottis position of C3 to C4 during phonation. The higher position of neoglottis upon phonation was driven by pulmonary air during exhalation. The alternation of the location of the source at the same time changes the vocal tract configuration in TE speakers. A higher position of neoglottis in TE speakers than the vocal folds in laryngeal speakers means a general reduction of effective length of vocal tract for resonance in TE speakers than in laryngeal speakers of Cantonese.

Fundamental frequency (F0) and myoelastic properties of PE segment

Results obtained from the EGG analysis of sustained phonation of all three vowels show that TE speakers have lower mean F0 than laryngeal speakers. This is coherent with the finding of lower F0 in TE speakers than laryngeal speakers in a recent study of EGG analysis of sustained /i/ production (Kazi et al., 2009a). As F0 was determined by length of vocal folds, tissue density and mean longitudinal stress of the source of vibration, the phenomenon of lower F0 could reasonably be the result of a larger total tissue mass of PE segment in TE speakers than that of vocal folds in laryngeal speakers. Moreover, a reduction in mean longitudinal stress could also result in a lower F0. These explanations of the lower F0 in TE speakers could possibly be the reasons for our results of present study, yet more research studies need to be carried out to verify the properties of PE segment in local TE speakers.

Closed Quotient and vibratory behaviors of PE segment

EGG analysis of mean F0 and contact quotient was found to be valid in examining the vibratory behaviors of laryngeal speakers of different age and gender groups (Ma, & Love, 2010). In this study, closed quotient in TE speakers were significantly greater than laryngeal speakers in both sustained vowel phonation and connected speech production. This implies a slower closing movement of the PE segment in TE speakers compared with the abrupt closure of vocal folds in laryngeal speakers during phonation. This implication could be, to certain extent, related to the higher tissue density of the PE segment than the vocal folds in laryngeal speakers, yet more researches are required to justify the hypothesis. In a study investigating anatomical and morphologic features of TE speakers using EGG-based videostroboscopy, Kazi et al. (2009b) found good TE speech is associated with floppy mucosa which has good volitional movement while poor TE speakers are featured with rigid mucosa. The finding was insightful and suggested that, besides tissue density, other anatomical or morphologic features of the PE segment should be investigated in order to make comprehensive hypothesis regarding the slow movement of the PE segment in TE speakers.

Suggestions for further studies

1. Future studies can employ multidimensional investigation of TE speech, for instance, the combination of acoustic, EGG and perceptual analysis to investigate the correlation between the objective parameters and perceptual judgment of voice quality.
2. There is a need to document the anatomical and morphologic features of the PE segment of local TE speakers which will provide vital knowledge about the tissue density, and other myoelastic properties of PE segment after total laryngectomy or subsequent pharyngeal reconstruction for the local population.
3. A more comprehensive investigation of the vibratory behaviors of PE segment in TE speakers by both quantitative and qualitative EGG analysis may yield a better understanding of the aperiodicity and quality of TE speech.

Conclusion

The literature reveals very limited studies investigating the differences between tracheoesophageal (TE) speech and laryngeal speech of Cantonese using both acoustic and electroglottography (EGG) analysis in both sustained vowel phonation and connected speech production. This study attempts to investigate the objective acoustic and EGG measures between the TE speakers and laryngeal speakers of Cantonese, infer their significances towards perceptual assessment and discuss the plausible explanation of the differences of measures between the two groups in sustained vowel phonation of /a/, /i/ and /u/, and passage reading.

In general, TE speakers had lower F0 for their perceptually low voice. It was explained in terms of reduced longitudinal stress and greater tissue density of the PE segment. They also had greater SD-F0, jitter and shimmer, and lower H/N which collectively contribute to perceptually hoarse and breathy voice. Formant frequencies (F1 & F2) were found to be higher in TE speakers than laryngeal speakers, reflecting a shorter effective length of resonance for TE speakers which confirm a higher position of sound source in TE phonation than normal laryngeal phonation. Moreover a higher closed quotient in TE speakers suggested a slower movement of PE segment mucosa during closing phase of its vibration cycle.

Acknowledgment

The author thanks Dr. Manwa L. Ng for his invaluable comments and support in this study. Special thanks are extended to Dr. Nan Yan for his kind support with using the instruments, and to the anonymous volunteers involved in this study. This study was supported by the New Voice Club of Hong Kong.

References

van As, C. J., Op de Coul, B. M. R., van den Hoogen, F. J. A., Koopmans-van Beinum, F. J., & Hilgers, F. J. M. (2001). Quantitative videofluoroscopy: a new evaluation tool for tracheoesophageal voice production. *Archives of Otolaryngology - Head & Neck Surgery, 127*, 161–169.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341-345.

Childers, D. G., Krishnamurthy, A. K., Naik, J. M, Larar, J. N., & Moore, G. P. (1983). Assessment of laryngeal function by simultaneous measurement of speech, electroglottography and ultra-high speed film. *Folia Phoniatrica et Logopaedia, 35,* 116.

Childers, D. G, Smith, A. M., & Moore, G. P. (1984). Relationships between electroglottograph, speech and vocal fold contact. *Folia Phoniatrica et Logopaedia, 36*, 105-118.

Fabre, P. (1957). Percutaneous electric process registering glottic union during phonation: glottography at high frequency; first results. *Bulletin de l Academie Nationale de Medecine (Paris), 141*, 66-69.

Kazi, R., Kanagalingam, J.,Venkitaraman, R., Prasad, V., Clarke, P., Nutting, C. M., Rhys-Evans, P., & Harrington, K. J. (2009a). Electroglottographic and perceptual evaluation of tracheoesophageal speech. *Journal of Voice, 23*(2), 247-254.

Kazi, R., Kiverniti, E., Prasad, V., Venkitaraman, R., Butting, C. M., Clarke, P., Rhys-Evans, P., & Harrington, K. J. (2006). Multidimensional assessment of female tracheoesophageal prosthetic speech. *Clinical Otolaryngology, 31*, 511-517.

Kazi, R. A., Singh, A., Venkitaraman, R., Sayed, S. I., Rhys-Evans, P., & Harrington, K. J. (2009b). Is electroglottography-based videostroboscopic assessment of post laryngectomy prosthetic speech useful? *Journal of Cancer Research and Therapeutics, 5,* 85-92.

Kitzing, P. (2000). Electroglottography. In A. Ferlito (Ed.), *Diseases of the Larynx* (pp. 127-138). London: Arnold.

Konstantopoulos, K., Vikelis, M., Seikel, J. A., & Mitsikostas, D. D. (2010). The existence of phonatory instability in multiple sclerosis: an acoustic and electroglottographic study. *Neurological Sciences, 31*, 259–268.

Liu, H., & Ng, M. (2009). Formant characteristics of vowels produced by Mandarin esophageal speakers. *Journal of Voice, 22* (2), 210-218.

Ma, E. P.-M., & Love, A. L. (2010). Electroglottographic Evaluation of Age and Gender Effects During Sustained Phonation and Connected Speech. *Journal of Voice, 24* (2), 146-152.

Most, T., Tobin, Y., & Mimran, R.C. (2000). Acoustic and perceptual characteristics of esophageal and tracheoesophageal speech production. *Journal of Communication Disorders, 33* (2), 165-181.

Ng, M., & Chu, R. (2009). An Acoustical and Perceptual Study of Vowels Produced by Alaryngeal Speakers of Cantonese. *Folia Phoniatrica et Logopaedia, 61*, 79-104.

Ng, M., Gilbert, H.R., & Lerman, J.W. (2001) Fundamental frequency, intensity, and vowel duration characteristics related to perception of Cantonese alaryngeal speech. *Folia Phoniatrica et Logopaedia, 53,* 36–47.

Ng, M., Kwok, C. L., & Chow, S. F. (1997). Speech performance of adult Cantonese-speaking laryngectomees using different types of alaryngeal phonation. *Journal of Voice, 11*(3), 338-344.

Ng, M., & Wong, J. (2009). Voice Onset Time Characteristics of Esophageal, Tracheoesophageal and Laryngeal Speech of Cantonese. *Journal of Speech, Language,* *and Hearing Research, 52*(3), 780-789.

Robbins, J., Fisher, H.B., Blom, E.C., & Mark, I.S. (1984). A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production. *Journal of Speech and Hearing Disorders, 49,* 202-210.

Rothenberg, M., & Mahshie, J. J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research, 31,* 338-351.

Singer, M.I., & Blom, E.D. (1980). An endoscopic technique for restoration of voice after laryngectomy. *Annals of Otology, Rhinology, and Laryngology, 89,* 529-533.

Sisty, N. L., & Weinberg, B. (1972). Formant frequency characteristics of esophageal speech. *Journal of Speech and Hearing Research, 15,* 439-448.

Stemple, J. C., Glaze, L. E., & Klaben, B. (2010). *Clinical Voice Pathology: Theory and Management*. San Diego, CA: Plural Publishing.

Appendix

巴巴拉史翠珊是我最喜愛的一位美國女明星，她沒有如花的美貌，但她出眾的技藝在「俏郎君」一劇中發揮得淋漓盡致，使這齣戲和它的主題曲都成爲傳頌一時的佳作。今天，銀圈裏的男男女女以色相自高，骨子裏只有自我宣傳的鬼主意，頻頻「車大炮」、造新聞，盼能提高知名度，而藝技高低，已與他們無關了。